

## **Leucine rich repeat containing protein**

This invention relates to a novel protein, termed INSP179, herein identified as a secreted protein, in particular, as a member of the leucine rich repeat containing protein family and to the use of this protein and nucleic acid sequence from the encoding gene in the  
5 diagnosis, prevention and treatment of disease.

All publications, patents and patent applications cited herein are incorporated in full by reference.

### **Background**

The process of drug discovery is presently undergoing a fundamental revolution as the era  
10 of functional genomics comes of age. The term "functional genomics" applies to an approach utilising bioinformatics tools to ascribe function to protein sequences of interest. Such tools are becoming increasingly necessary as the speed of generation of sequence data is rapidly outpacing the ability of research laboratories to assign functions to these protein sequences.

15 As bioinformatics tools increase in potency and in accuracy, these tools are rapidly replacing the conventional techniques of biochemical characterisation. Indeed, the advanced bioinformatics tools used in identifying the present invention are now capable of outputting results in which a high degree of confidence can be placed.

Various institutions and commercial organisations are examining sequence data as they  
20 become available and significant discoveries are being made on an on-going basis. However, there remains a continuing need to identify and characterise further genes and the polypeptides that they encode, as targets for research and for drug discovery.

### **Introduction**

#### **Secreted Proteins**

25 The ability for cells to make and secrete extracellular proteins is central to many biological processes. Enzymes, growth factors, extracellular matrix proteins and signalling molecules are all secreted by cells. This is through fusion of a secretory vesicle with the plasma membrane. In most cases, but not all, proteins are directed to the endoplasmic reticulum and into secretory vesicles by a signal peptide. Signal peptides are cis-acting sequences  
30 that affect the transport of polypeptide chains from the cytoplasm to a membrane bound

compartment such as a secretory vesicle. Polypeptides that are targeted to the secretory vesicles are either secreted into the extracellular matrix or are retained in the plasma membrane. The polypeptides that are retained in the plasma membrane will have one or more transmembrane domains. Examples of secreted proteins that play a central role in the functioning of a cell are cytokines, hormones, extracellular matrix proteins (adhesion molecules), proteases, and growth and differentiation factors. Description of some of the properties of these proteins follows.

### Introduction to Leucine-rich repeat domains

The Leucine-rich repeat (LRR) motif is characterised by tandem arrays of a leucine-rich consensus sequence and a structurally conserved LRR interaction surface that mediate reversible, high affinity protein-protein and protein-RNA interactions (Iozzo, R. V., 1998 *Crit. Rev. Biochem. Mol. Biol.* 32, 141-174; Kajava, A. V. 1998, *J. Mol. Biol.* 277, 519-527; Kobe, B. & Deisenhofer, J. 1995, *Curr. Opin. Struct. Biol.* 5, 409-416). More than 100 LRR-containing proteins have been identified from a diversity of eukaryotic organisms. At least six subfamilies are recognised based upon the differing lengths and consensus sequences of the repeats; 'typical' repeats consist of 20-27 residues. X-ray structural data are currently available for the LRR domains of several proteins including porcine and human placental ribonuclease inhibitor (RI), *Schizosaccharomyces pombe* rna1p, the RNA-binding human spliceosomal U2A' the internalin B protein of the bacterium *Listeria monocytogenes*, and the human mRNA export factor TAP. Although the number of tandem repeats in these LRR structures is variable, ranging from 5 to 17 repeats, the overall non-globular topology of each LRR is strongly similar. Each repeat is a structural unit consisting of a highly conserved  $\beta$  strand packed against a more variable strand, usually helical; these are assembled along a common axis into an arc-shaped structure lined with parallel  $\beta$  strands along the inner surface, an adjoining asparagine or cysteine loop region, and the  $\alpha$ -helix, 310 helix, or extended conformation flanking the outer circumference. The solvent exposed parallel  $\beta$ -sheet lining the inner cleft and the curvature of the structure are general features of the LRR motif. Several studies point to the non-leucine, interstitial residues of the solvent exposed  $\beta$ -sheet and the  $\beta/\alpha$  turn regions of the LRR motif as involved in protein and RNA ligand recognition and binding (22, 26, 30-32). However, as each of these studies has targeted only a few of the potential LRR contact points, it is not yet clear whether such limited interactions are universally employed by this diverse protein family to mediate binding (Clark, L.B. *et al.* 2004, JBC).

These domains are found in a number of proteins with diverse functions such as immune response, hormone-receptor interactions, enzyme inhibition, vascular repair, cell adhesion and cellular trafficking. Studies have shown the involvement of LRR proteins in early mammalian development, neural development, cell polarization, regulation of gene expression and apoptosis signalling. LRRs may also be critical to the morphology and dynamics of cytoskeleton. Examples of such proteins include CD180 antigen precursor (Lymphocyte antigen 64), MHC class II transactivator (CIITA), Toll-like receptor precursor, Follicle stimulating hormone receptor precursor (FSH-R) (Follitropin receptor), Lutropin-choriogonadotropic hormone receptor precursor, Leucine-rich repeat-containing G protein-coupled receptor 4 precursor, Platelet glycoprotein IX precursor (GPIX) (CD42A), Oligodendrocyte-myelin glycoprotein precursor, Photoreceptor-associated LRR superfamily protein precursor (Retina specific protein PAL) and High affinity nerve growth factor receptor precursor. Interestingly, certain proteins only seem to contain leucine rich repeat domains, such as, Chondroadherin precursor (cartilage formation), Nogo 66 receptor (inhibition of axon growth) and CD14 (immune response).

Thus there is a great need for the identification of novel leucine rich repeat containing proteins, as these proteins are highly likely to be implicated in disease. Identification of such proteins will also be of importance in increasing the understanding of the underlying pathways that lead to the diseases states and associated disease states, mentioned above, and in developing more effective gene and/or drug therapies to treat these disorders.

## THE INVENTION

The invention is based on the discovery that the INSP179 polypeptide is a member of the Leucine rich repeat containing protein family.

In one embodiment of the first aspect of the invention, there is provided a polypeptide which:

- (i) comprises the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and/or SEQ ID NO:14;
- (ii) is a fragment thereof which is a member of the leucine rich repeat containing protein family, or having an antigenic determinant in common with the polypeptides of (i); or

(iii) is a functional equivalent of (i) or (ii).

Preferably, the polypeptide according to this first aspect of the invention comprises the amino acid sequence as recited in SEQ ID NO:10 or SEQ ID NO:14.

According to a second embodiment of this first aspect of the invention, there is provided a  
5 polypeptide which consists of the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO:18, SEQ ID NO: 20 and/or SEQ ID NO: 22.

The polypeptides consisting of the amino acids sequences recited in SEQ ID NO: 16 and SEQ ID NO: 18 are particularly preferred.

10 The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as “the INSP179 exon 1 polypeptide”. The polypeptide having the sequence recited in SEQ ID NO:4 is referred to hereafter as “the INSP179 exon 2 polypeptide”. The polypeptide having the sequence recited in SEQ ID NO:6 is referred to hereafter as “the INSP179 exon 3 polypeptide”. The polypeptide having the sequence recited in SEQ ID NO:8 is referred to  
15 hereafter as “the INSP179 exon 4 polypeptide”. The polypeptide having the sequence recited in SEQ ID NO:10 is referred to hereafter as “the INSP179 polypeptide”. The polypeptide having the sequence recited in SEQ ID NO:16 is referred to hereafter as “the INSP179-EC polypeptide”. The polypeptide having the sequence recited in SEQ ID NO:18 is referred to hereafter as “the INSP179-EC-SV1 polypeptide”.

20 The INSP179-EC polypeptide comprises all of exons 1, 2, 3 and the first 78 amino acids of exon 4 from the INSP179 polypeptide. The INSP179-EC-SV1 polypeptide comprises all of exon 1 and the first 78 amino acids of exon 4 from the INSP179 polypeptide.

Although the Applicant does not wish to be bound by this theory, it is postulated that the first 24 amino acids of the INSP179 polypeptide form a signal peptide.

25 The INSP179 exon 1 polypeptide without this postulated signal sequence is recited in SEQ ID NO: 12. The full length INSP179 polypeptide sequence without this postulated signal sequence is recited in SEQ ID NO: 14. The INSP179-EC polypeptide without this postulated signal sequence is recited in SEQ ID NO: 20. The INSP179-EC-SV1 polypeptide without this postulated signal sequence is recited in SEQ ID NO: 22.

30 The polypeptide having the sequence recited in SEQ ID NO: 12 is referred to hereafter as “the INSP179 exon 1 mature polypeptide”. The polypeptide having the sequence recited in

SEQ ID NO: 14 is referred to hereafter as “the INSP179 mature polypeptide”. The polypeptide having the sequence recited in SEQ ID NO: 20 is referred to hereafter as “the INSP179-EC mature polypeptide”. The polypeptide having the sequence recited in SEQ ID NO: 22 is referred to hereafter as “the INSP179-EC-SV1 mature polypeptide”.

- 5 The polypeptides of the first aspect of the invention may further comprise a histidine tag. Preferably the histidine tag is found at the C-terminal of the polypeptide. Preferably the histidine tag comprises 1-10 histidine residues (*e.g.* 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 residues). More preferably, the histidine tag comprises 6 histidine residues.

- Preferred amino acid sequences comprising histidine tags are recited in SEQ ID NO: 24,  
10 SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32 and SEQ ID NO: 34.

- The polypeptide having the sequence recited in SEQ ID NO: 24 is referred to hereafter as “the histidine tagged INSP179 polypeptide”. The polypeptide having the sequence recited in SEQ ID NO: 26 is referred to hereafter as “the histidine tagged INSP179 mature polypeptide”. The polypeptide having the sequence recited in SEQ ID NO: 28 is referred to  
15 hereafter as “the histidine tagged INSP179-EC polypeptide”. The polypeptide having the sequence recited in SEQ ID NO: 30 is referred to hereafter as “the histidine tagged INSP179-EC-SV1 polypeptide”. The polypeptide having the sequence recited in SEQ ID NO: 32 is referred to hereafter as “the histidine tagged INSP179-EC mature polypeptide”. The polypeptide having the sequence recited in SEQ ID NO: 34 is referred to hereafter as  
20 “the histidine tagged INSP179-EC-SV1 mature polypeptide”.

- The term “INSP179 polypeptides” as used herein includes the INSP179 exon 1 polypeptide, the INSP179 exon 2 polypeptide, the INSP179 exon 3 polypeptide, the INSP179 exon 4 polypeptide, the INSP179 polypeptide, the INSP179 exon 1 mature polypeptide, the INSP179 mature polypeptide, the INSP179-EC polypeptide, the INSP179-  
25 EC-SV1 polypeptide, the INSP179-EC mature polypeptide, the INSP179-EC-SV1 mature polypeptide, the histidine tagged INSP179 polypeptide, the histidine tagged INSP179 mature polypeptide, the histidine tagged INSP179-EC polypeptide, the histidine tagged INSP179-EC-SV1 polypeptide, the histidine tagged INSP179-EC mature polypeptide and the histidine tagged INSP179-EC-SV1 mature polypeptide.

- 30 Preferably, a polypeptide according to any one of the above-described aspects of the invention functions as a member of the leucine rich repeat containing family of proteins.

By “functions as a member of the leucine rich repeat containing family of proteins” we refer to polypeptides that comprise amino acid sequence or structural features that can be identified as conserved features within the polypeptides of the leucine rich repeat containing family of proteins. Such proteins are likely to play a role in the cell signalling process as either receptors or ligands. Upregulation of the expression of this protein by bacterial lipopolysaccharide (LPS) suggests a role in the pro-inflammatory response. Therefore polypeptides having a similar function to INSP179 may be detected by treating cells with LPS and detecting whether expression of that protein is upregulated.

Leucine rich repeat containing proteins such as CD180 antigen precursor (Lymphocyte antigen 64), MHC class II transactivator (CIITA), Toll-like receptor 6 precursor and CD14 are known to play a role in the immune response. LPS is often used to induce a pro-inflammatory response in immune system cells and is known to induce the production of cytokines and inflammatory mediators such as TNF, IL-1, IL-6 and IL-8. By showing that INSP179 expression may be upregulated by treating T cells such as THP-1 cells with LPS *in vitro*, a link can be made between the known LRR containing proteins that play a role in the immune response and the INSP179 polypeptide.

Preferably a polypeptide as described above is glycosylated. The INSP179 polypeptide (SEQ ID NO: 10) may be glycosylated at one or more sites selected from amino acid 45, 115, 454, 746 and 756. It will be appreciated that fragments and/or functional equivalents of the INSP179 polypeptide may be glycosylated at the same sites, though the amino acid numbering may be different due to insertions/deletions. Such glycosylation is indicative of extracellular proteins.

In a second aspect, the invention provides a purified nucleic acid molecule which encodes a polypeptide of the first aspect of the invention.

Preferably, the purified nucleic acid molecule comprises the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the INSP179 exon 1 polypeptide), SEQ ID NO:3 (encoding the INSP179 exon 2 polypeptide), SEQ ID NO:5 (encoding the INSP179 exon 3 polypeptide), SEQ ID NO:7 (encoding the INSP179 exon 4 polypeptide), SEQ ID NO:9 (encoding the INSP179 polypeptide), SEQ ID NO:11 (encoding the INSP179 mature exon 1 polypeptide), SEQ ID NO:13 (encoding the INSP179 mature polypeptide), SEQ ID NO:15 (encoding the INSP179-EC polypeptide), SEQ ID NO:17 (encoding the INSP179-EC-SV1 polypeptide), SEQ ID NO:19 (encoding the INSP179-EC mature polypeptide), SEQ

ID NO: 21 (encoding the INSP179-EC-SV1 mature polypeptide), SEQ ID NO: 23 (encoding the histidine tagged INSP179 polypeptide), SEQ ID NO: 25 (encoding the histidine tagged INSP179 mature polypeptide), SEQ ID NO: 27 (encoding the histidine tagged INSP179-EC polypeptide), SEQ ID NO: 29 (encoding the histidine tagged  
5 INSP179-EC-SV1 polypeptide), SEQ ID NO: 31 (encoding the histidine tagged INSP179-EC mature polypeptide) and/or SEQ ID NO: 33 (encoding the histidine tagged INSP179-EC-SV1 polypeptide).

The invention further provides that the purified nucleic acid molecule consists of the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the INSP179 exon 1  
10 polypeptide), SEQ ID NO:3 (encoding the INSP179 exon 2 polypeptide), SEQ ID NO:5 (encoding the INSP179 exon 3 polypeptide), SEQ ID NO:7 (encoding the INSP179 exon 4 polypeptide), SEQ ID NO:9 (encoding the INSP179 polypeptide), SEQ ID NO:11 (encoding the INSP179 mature exon 1 polypeptide), SEQ ID NO:13 (encoding the INSP179 mature polypeptide), SEQ ID NO: 15 (encoding the INSP179-EC polypeptide),  
15 SEQ ID NO: 17 (encoding the INSP179-EC-SV1 polypeptide), SEQ ID NO: 19 (encoding the INSP179-EC mature polypeptide), SEQ ID NO: 21 (encoding the INSP179-EC-SV1 mature polypeptide), SEQ ID NO: 23 (encoding the histidine tagged INSP179 polypeptide), SEQ ID NO: 25 (encoding the histidine tagged INSP179 mature polypeptide), SEQ ID NO: 27 (encoding the histidine tagged INSP179-EC polypeptide),  
20 SEQ ID NO: 29 (encoding the histidine tagged INSP179-EC-SV1 polypeptide), SEQ ID NO: 31 (encoding the histidine tagged INSP179-EC mature polypeptide) and/or SEQ ID NO: 33 (encoding the histidine tagged INSP179-EC-SV1 polypeptide).

In a third aspect, the invention provides a purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule of the second aspect of the  
25 invention.

In a fourth aspect, the invention provides a vector, such as an expression vector, that contains a nucleic acid molecule of the second or third aspect of the invention.

In a fifth aspect, the invention provides a host cell transformed with a vector of the fourth aspect of the invention.

30 In a sixth aspect, the invention provides a ligand which binds specifically to protein members of the leucine rich repeat containing family of proteins of the first aspect of the invention. Preferably, the ligand inhibits the function of a polypeptide of the first aspect of

the invention which is a member of the leucine rich repeat containing family of proteins. Ligands to a polypeptide according to the invention may come in various forms, including natural or modified substrates, enzymes, receptors, small organic molecules such as small natural or synthetic organic molecules of up to 2000Da, preferably 800Da or less, 5 peptidomimetics, inorganic molecules, peptides, polypeptides, antibodies, structural or functional mimetics of the aforementioned.

In a seventh aspect, the invention provides a compound that is effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

10 A compound of the seventh aspect of the invention may either increase (agonise) or decrease (antagonise) the level of expression of the gene or the activity of the polypeptide.

Importantly, the identification of the function of the INSP179 polypeptides allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of disease. Ligands and compounds according to the sixth and 15 seventh aspects of the invention may be identified using such methods. These methods are included as aspects of the present invention.

In an eighth aspect, the invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the 20 invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in therapy or diagnosis of diseases in which members of the leucine rich repeat containing family of proteins are implicated. Such diseases include cell proliferative disorders, autoimmune/inflammatory disorders, such as arthritis, cardiovascular disorders, neurological disorders, developmental disorders, fertility 25 disorders, metabolic disorders, AIDS, renal disease, infections and other pathological conditions, particularly those in which leucine rich repeat containing proteins are implicated. These molecules may also be used in the manufacture of a medicament for the treatment such diseases. Leucine rich repeat containing proteins may also be useful in designing new methods of contraception, as will compounds that bind to these proteins.

30 In a ninth aspect, the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide of the first aspect of the invention or the activity of a polypeptide of the first aspect of the



invention in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease. Such a method will preferably be carried out *in vitro*. Similar methods may be used for monitoring the therapeutic treatment of disease in a patient, wherein altering the level of  
5 expression or activity of a polypeptide or nucleic acid molecule over the period of time towards a control level is indicative of regression of disease.

A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand, such as an antibody, of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a  
10 ligand-polypeptide complex; and (b) detecting said complex.

A number of different such methods according to the ninth aspect of the invention exist, as the skilled reader will be aware, such as methods of nucleic acid hybridization with short probes, point mutation analysis, polymerase chain reaction (PCR) amplification and methods using antibodies to detect aberrant protein levels. Similar methods may be used  
15 on a short or long term basis to allow therapeutic treatment of a disease to be monitored in a patient. The invention also provides kits that are useful in these methods for diagnosing disease.

In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of the invention as a leucine rich repeat containing protein, such as in cell communication and  
20 development of cells.

In an eleventh aspect, the invention provides a pharmaceutical composition comprising a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a  
25 compound of the seventh aspect of the invention, in conjunction with a pharmaceutically-acceptable carrier.

In a twelfth aspect, the present invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the  
30 invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in the manufacture of a medicament for the diagnosis or treatment of a disease.

In a thirteenth aspect, the invention provides a method of treating a disease in a patient comprising administering to the patient a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand  
5 of the sixth aspect of the invention, or a compound of the seventh aspect of the invention.

For diseases in which the expression of a natural gene encoding a polypeptide of the first aspect of the invention, or in which the activity of a polypeptide of the first aspect of the invention, is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound  
10 administered to the patient should be an agonist. Conversely, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an antagonist. Examples of such antagonists include antisense nucleic acid molecules,  
15 ribozymes and ligands, such as antibodies.

The INSP179 polypeptides are members of the leucine rich repeat containing family of proteins and thus have roles in many disease states. Antagonists of the INSP179 polypeptides are of particular interest as they provide a way of modulating these disease states.

20 In a fourteenth aspect, the invention provides transgenic or knockout non-human animals that have been transformed to express higher, lower or absent levels of a polypeptide of the first aspect of the invention. Such transgenic animals are very useful models for the study of disease and may also be used in screening regimes for the identification of compounds that are effective in the treatment or diagnosis of such a disease.

25 A summary of standard techniques and procedures which may be employed in order to utilise the invention is given below. It will be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and it is not intended that this terminology should limit the  
30 scope of the present invention. The extent of the invention is limited only by the terms of the appended claims.

Standard abbreviations for nucleotides and amino acids are used in this specification.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA technology and immunology, which are within the skill of those working in the art.

Such techniques are explained fully in the literature. Examples of particularly suitable texts for consultation include the following: Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 & 155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Second Edition (Springer Verlag, N.Y.); and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds. 1986).

As used herein, the term "polypeptide" includes any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.* peptide isosteres. This term refers both to short chains (peptides and oligopeptides) and to longer chains (proteins).

The polypeptide of the present invention may be in the form of a mature protein or may be a pre-, pro- or prepro- protein that can be activated by cleavage of the pre-, pro- or prepro- portion to produce an active mature polypeptide. In such polypeptides, the pre-, pro- or prepro- sequence may be a leader or secretory sequence or may be a sequence that is employed for purification of the mature polypeptide sequence.

The polypeptide of the first aspect of the invention may form part of a fusion protein. For example, it is often advantageous to include one or more additional amino acid sequences which may contain secretory or leader sequences, pro-sequences, sequences which aid in purification, or sequences that confer higher protein stability, for example during recombinant production. Alternatively or additionally, the mature polypeptide may be fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Preferably, the polypeptide of the invention comprising a sequence having homology with INSP179, INSP179-EC or INSP179-EC-SV1 is a fusion protein. Such fusion proteins can be obtained by cloning a polynucleotide encoding a polypeptide comprising a sequence having at least 85% homology INSP179, INSP179-EC or INSP179-EC-SV1 in frame with  
5 the coding sequences for a heterologous protein sequence.

The term "heterologous", when used herein, is intended to designate any polypeptide other than a human INSP179, INSP179-EC or INSP179-EC-SV1 polypeptide. Examples of heterologous sequences, that can be comprised in the fusion proteins either at the N- or C-terminus, include: extracellular domains of membrane-bound protein, immunoglobulin  
10 constant regions (Fc regions), multimerization domains, domains of extracellular proteins, signal sequences, export sequences, and sequences allowing purification by affinity chromatography.

Many of these heterologous sequences are commercially available in expression plasmids since these sequences are commonly included in fusion proteins in order to provide  
15 additional properties without significantly impairing the specific biological activity of the protein fused to them (Terpe K, 2003, Appl Microbiol Biotechnol, 60:523-33). Examples of such additional properties are a longer lasting half-life in body fluids, the extracellular localization, or an easier purification procedure as allowed by the a stretch of Histidines forming the so-called "histidine tag" (Gentz *et al.* 1989, Proc Natl Acad Sci USA, 86:821-  
20 4) or by the "HA" tag, an epitope derived from the influenza hemagglutinin protein (Wilson *et al.* 1994, Cell, 37:767-78). If needed, the heterologous sequence can be eliminated by a proteolytic cleavage, for example by inserting a proteolytic cleavage site between the protein and the heterologous sequence, and exposing the purified fusion protein to the appropriate protease. These features are of particular importance for the  
25 fusion proteins since they facilitate their production and use in the preparation of pharmaceutical compositions. For example, the INSP179, INSP179-EC or INSP179-EC-SV1 polypeptide may be purified by means of a hexa-histidine peptide fused at the C-terminus of INSP179, INSP179-EC or INSP179-EC-SV1. When the fusion protein comprises an immunoglobulin region, the fusion may be direct, or via a short linker  
30 peptide which can be as short as 1 to 3 amino acid residues in length or longer, for example, 13 amino acid residues in length. Said linker may be a tripeptide of the sequence E-F-M (Glu-Phe-Met), for example, or a 13-amino acid linker sequence comprising Glu-Phe-Gly-Ala-Gly-Leu-Val-Leu-Gly-Gly-Gln-Phe-Met (SEQ ID NO: 35) introduced

between the sequence of the substances of the invention and the immunoglobulin sequence. The resulting fusion protein has improved properties, such as an extended residence time in body fluids (*i.e.* an increased half-life), increased specific activity, increased expression level, or the purification of the fusion protein is facilitated.

- 5 In a preferred embodiment, the protein is fused to the constant region of an Ig molecule. Preferably, it is fused to heavy chain regions, like the CH2 and CH3 domains of human IgG1, for example. Other isoforms of Ig molecules are also suitable for the generation of fusion proteins according to the present invention, such as isoforms IgG2 or IgG4, or other Ig classes, like IgM or IgA, for example. Fusion proteins may be monomeric or  
10 multimeric, hetero- or homomultimeric.

In a further preferred embodiment, the functional derivative comprises at least one moiety attached to one or more functional groups, which occur as one or more side chains on the amino acid residues. Preferably, the moiety is a polyethylene (PEG) moiety. PEGylation may be carried out by known methods, such as the ones described in WO99/55377, for  
15 example.

- Polypeptides may contain amino acids other than the 20 gene-encoded amino acids, modified either by natural processes, such as by post-translational processing or by chemical modification techniques which are well known in the art. Among the known modifications which may commonly be present in polypeptides of the present invention  
20 are glycosylation, lipid attachment, sulphation, gamma-carboxylation, for instance of glutamic acid residues, hydroxylation and ADP-ribosylation. Other potential modifications include acetylation, acylation, amidation, covalent attachment of flavin, covalent attachment of a haeme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid derivative, covalent attachment of  
25 phosphatidylinositol, cross-linking, cyclization, disulphide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, GPI anchor formation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.
- 30 Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl terminus in a polypeptide, or both, by a covalent modification is common in

naturally-occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention.

The modifications that occur in a polypeptide often will be a function of how the polypeptide is made. For polypeptides that are made recombinantly, the nature and extent  
5 of the modifications in large part will be determined by the post-translational modification capacity of the particular host cell and the modification signals that are present in the amino acid sequence of the polypeptide in question. For instance, glycosylation patterns vary between different types of host cell.

The polypeptides of the present invention can be prepared in any suitable manner. Such  
10 polypeptides include isolated naturally-occurring polypeptides (for example purified from cell culture), recombinantly-produced polypeptides (including fusion proteins), synthetically-produced polypeptides or polypeptides that are produced by a combination of these methods.

The functionally-equivalent polypeptides of the first aspect of the invention may be  
15 polypeptides that are homologous to the INSP179 polypeptides. Two polypeptides are said to be "homologous", as the term is used herein, if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid residue is identical between the sequences. "Similarity" indicates that, at any particular  
20 position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G.,  
25 eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information;  
30 <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

Homologous polypeptides therefore include natural biological variants (for example,

allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the INSP179 polypeptides. Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr. Particularly preferred are variants in which several, *i.e.* between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions. Such mutants also include polypeptides in which one or more of the amino acid residues includes a substituent group.

Typically, greater than 30% identity between two polypeptides is considered to be an indication of functional equivalence. Preferably, functionally equivalent polypeptides of the first aspect of the invention have a degree of sequence identity with the INSP179 polypeptide, or with active fragments thereof, of greater than 80%. More preferred polypeptides have degrees of identity of greater than 85%, 90%, 95%, 98% or 99%, respectively.

The functionally-equivalent polypeptides of the first aspect of the invention may also be polypeptides which have been identified using one or more techniques of structural alignment. For example, the Inpharmatica Genome Threader technology that forms one aspect of the search tools used to generate the Biopendium™ search database may be used (see PCT application WO 01/69507) to identify polypeptides of presently-unknown function which, while having low sequence identity as compared to the INSP179 polypeptides, are predicted to be members of the leucine rich repeat containing family of proteins, by virtue of sharing significant structural homology with the INSP179 polypeptide sequence. By "significant structural homology" is meant that the Inpharmatica Genome Threader predicts two proteins to share structural homology with a certainty of 10% and above.

Polypeptides may be divided into fragments and similarly fragments of functional equivalents may exist. Such fragments are identified by being members of the same protein family as the full-length polypeptide, or having an antigenic determinant in common with the full-length polypeptides.

- 5 As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of a polypeptide or one of the functional equivalents of that polypeptide. The fragments should comprise at least *n* consecutive amino acids from the sequence and, depending on the particular sequence, *n* preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small fragments  
10 may form an antigenic determinant.

Fragments of full length polypeptides may consist of combinations of 1, 2, 3 or all 4 neighbouring exon sequences in the polypeptide sequences, respectively. For example, such combinations include exons 1 and 2, exons 2 and 3 or exons 1, 2 and 3, and so on. Such fragments are included in the present invention.

- 15 Such fragments may be "free-standing", *i.e.* not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the fragment of the invention most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre- and/or pro- polypeptide region fused to the  
20 amino terminus of the fragment and/or an additional region fused to the carboxyl terminus of the fragment. However, several fragments may be comprised within a single larger polypeptide.

- The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or  
25 monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones expressing the polypeptides of the invention or to purify the polypeptides by affinity chromatography. The antibodies may also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

- 30 The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to



fragments thereof, such as Fab, F(ab')<sub>2</sub> and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the first aspect of the invention.

By "substantially greater affinity" we mean that there is a measurable increase in the affinity for a polypeptide of the invention as compared with the affinity for known secreted proteins.

Preferably, the affinity is at least 1.5-fold, 2-fold, 5-fold 10-fold, 100-fold, 10<sup>3</sup>-fold, 10<sup>4</sup>-fold, 10<sup>5</sup>-fold, 10<sup>6</sup>-fold or greater for a polypeptide of the invention than for known secreted proteins such as members of the leucine rich repeat containing family of proteins.

10 If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The polypeptide used to immunise the animal can be derived by recombinant DNA technology or can be synthesized chemically. If desired, the polypeptide can be conjugated to a carrier protein. Commonly used carriers to which the polypeptides may be chemically coupled  
15 include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be  
20 readily produced by one skilled in the art. The general methodology for making monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, 77-96 in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985).

25 Panels of monoclonal antibodies produced against the polypeptides of the first aspect of the invention can be screened for various properties, *i.e.*, for isotype, epitope, affinity, etc. Monoclonal antibodies are particularly useful in purification of the individual polypeptides against which they are directed. Alternatively, genes encoding the monoclonal antibodies of interest may be isolated from hybridomas, for instance by PCR techniques known in the  
30 art, and cloned and expressed in appropriate vectors.

Chimeric antibodies, in which non-human variable regions are joined or fused to human

constant regions (see, for example, Liu *et al.*, Proc. Natl. Acad. Sci. USA, 84, 3439 (1987)), may also be of use.

The antibody may be modified to make it less immunogenic in an individual, for example by humanisation (see Jones *et al.*, Nature, 321, 522 (1986); Verhoeyen *et al.*, Science, 239, 1534 (1988); Kabat *et al.*, J. Immunol., 147, 1709 (1991); Queen *et al.*, Proc. Natl. Acad. Sci. USA, 86, 10029 (1989); Gorman *et al.*, Proc. Natl Acad. Sci. USA, 88, 34181 (1991); and Hodgson *et al.*, Bio/Technology, 9, 421 (1991)). The term "humanised antibody", as used herein, refers to antibody molecules in which the CDR amino acids and selected other amino acids in the variable domains of the heavy and/or light chains of a non-human donor antibody have been substituted in place of the equivalent amino acids in a human antibody. The humanised antibody thus closely resembles a human antibody but has the binding ability of the donor antibody.

In a further alternative, the antibody may be a "bispecific" antibody, that is, an antibody having two different antigen binding domains, each domain being directed against a different epitope.

Phage display technology may be utilised to select genes which encode antibodies with binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the relevant antibodies, or from naive libraries (McCafferty, J. *et al.*, (1990), Nature 348, 552-554; Marks, J. *et al.*, (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) Nature 352, 624-628).

Antibodies generated by the above techniques, whether polyclonal or monoclonal, have additional utility in that they may be employed as reagents in immunoassays, radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA). In these applications, the antibodies can be labelled with an analytically-detectable reagent such as a radioisotope, a fluorescent molecule or an enzyme.

Preferred nucleic acid molecules of the second and third aspects of the invention are those which encode a polypeptide sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 and SEQ ID NO:34 and functionally equivalent polypeptides. These nucleic acid molecules may be used in the methods and applications

described herein. The nucleic acid molecules of the invention preferably comprise at least n consecutive nucleotides from the sequences disclosed herein where, depending on the particular sequence, n is 10 or more (for example, 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

- 5 The nucleic acid molecules of the invention also include sequences that are complementary to nucleic acid molecules described above (for example, for antisense or probing purposes).

Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance cDNA, synthetic DNA or genomic  
10 DNA. Such nucleic acid molecules may be obtained by cloning, by chemical synthetic techniques or by a combination thereof. The nucleic acid molecules can be prepared, for example, by chemical synthesis using techniques such as solid phase phosphoramidite chemical synthesis, from genomic or cDNA libraries or by separation from an organism. RNA molecules may generally be generated by the *in vitro* or *in vivo* transcription of DNA  
15 sequences.

The nucleic acid molecules may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The term "nucleic acid molecule" also includes analogues of DNA and RNA, such as those  
20 containing modified backbones, and peptide nucleic acids (PNA). The term "PNA", as used herein, refers to an antisense molecule or an anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, which preferably ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell, where they  
25 preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. *et al.* (1993) *Anticancer Drug Des.* 8:53-63).

A nucleic acid molecule which encodes a polypeptide of this invention may be identical to the coding sequence of one or more of the nucleic acid molecules disclosed herein.

These molecules also may have a different sequence which, as a result of the degeneracy  
30 of the genetic code, encodes a polypeptide SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID

NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 or SEQ ID NO: 34. Such nucleic acid molecules may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those  
5 encoding a leader or secretory sequence, such as a pro-, pre- or prepro- polypeptide sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals),  
10 ribosome binding and mRNA stability. The nucleic acid molecules may also include additional sequences which encode additional amino acids, such as those which provide additional functionalities.

The nucleic acid molecules of the second and third aspects of the invention may also encode the functional equivalents of the polypeptides of the first aspect of the invention.  
15 Such a nucleic acid molecule may be a naturally-occurring variant such as a naturally-occurring allelic variant, or the molecule may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells or organisms.

20 Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or insertions. The substitutions, deletions or insertions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or insertions.

25 The nucleic acid molecules of the invention can also be engineered, using methods generally known in the art, for a variety of reasons, including modifying the cloning, processing, and/or expression of the gene product (the polypeptide). DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides are included as techniques which may be used to engineer the nucleotide  
30 sequences. Site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations and so forth.

Nucleic acid molecules which encode a polypeptide of the first aspect of the invention may be ligated to a heterologous sequence so that the combined nucleic acid molecule encodes a fusion protein. Such combined nucleic acid molecules are included within the second or third aspects of the invention. For example, to screen peptide libraries for inhibitors of the  
5 activity of the polypeptide, it may be useful to express, using such a combined nucleic acid molecule, a fusion protein that can be recognised by a commercially-available antibody. A fusion protein may also be engineered to contain a cleavage site located between the sequence of the polypeptide of the invention and the sequence of a heterologous protein so that the polypeptide may be cleaved and purified away from the heterologous protein.

10 The nucleic acid molecules of the invention also include antisense molecules that are partially complementary to nucleic acid molecules encoding polypeptides of the present invention and that therefore hybridize to the encoding nucleic acid molecules (hybridization). Such antisense molecules, such as oligonucleotides, can be designed to recognise, specifically bind to and prevent transcription of a target nucleic acid encoding a  
15 polypeptide of the invention, as will be known by those of ordinary skill in the art (see, for example, Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee *et al.*, Nucleic Acids Res 6, 3073 (1979); Cooney *et al.*, Science 241, 456 (1988); Dervan *et al.*, Science 251, 1360 (1991).

The term "hybridization" as used here refers to the association of two nucleic acid  
20 molecules with one another by hydrogen bonding. Typically, one molecule will be fixed to a solid support and the other will be free in solution. Then, the two molecules may be placed in contact with one another under conditions that favour hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase  
25 molecule to the solid support (Denhardt's reagent or BLOTTO); the concentration of the molecules; use of compounds to increase the rate of association of molecules (dextran sulphate or polyethylene glycol); and the stringency of the washing conditions following hybridization (see Sambrook *et al.* [*supra*]).

The inhibition of hybridization of a completely complementary molecule to a target  
30 molecule may be examined using a hybridization assay, as known in the art (see, for example, Sambrook *et al.* [*supra*]). A substantially homologous molecule will then compete for and inhibit the binding of a completely homologous molecule to the target

molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511).

"Stringency" refers to conditions in a hybridization reaction that favour the association of  
5 very similar molecules over association of molecules that differ. High stringency hybridisation conditions are defined as overnight incubation at 42°C in a solution comprising 50% formamide, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardts solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in  
10 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation reaction being carried out at 35°C (see Sambrook *et al.* [*supra*]). Preferably, the conditions used for hybridization are those of high stringency.

Preferred embodiments of this aspect of the invention are nucleic acid molecules that are at least 70% identical over their entire length to a nucleic acid molecule encoding the  
15 INSP179 polypeptides and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to such coding sequences, or is a nucleic acid molecule that is complementary thereto. In this regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at  
20 least 98%, 99% or more identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides which retain substantially the same biological function or activity as the INSP179 polypeptides.

The invention also provides a process for detecting a nucleic acid molecule of the  
25 invention, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting any such duplexes that are formed.

As discussed additionally below in connection with assays that may be utilised according to the invention, a nucleic acid molecule as described above may be used as a  
30 hybridization probe for RNA, cDNA or genomic DNA, in order to isolate full-length cDNAs and genomic clones encoding the INSP179 polypeptides and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity

to the gene encoding this polypeptide.

In this regard, the following techniques, among others known in the art, may be utilised and are discussed below for purposes of illustration. Methods for DNA sequencing and analysis are well known and are generally available in the art and may, indeed, be used to  
5 practice many of the embodiments of the invention discussed herein. Such methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proof-reading exonucleases such as those found in the ELONGASE Amplification System marketed by  
10 Gibco/BRL (Gaithersburg, MD). Preferably, the sequencing process may be automated using machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), the Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

One method for isolating a nucleic acid molecule encoding a polypeptide with an  
15 equivalent function to that of the INSP179 polypeptide is to probe a genomic or cDNA library with a natural or artificially-designed probe using standard procedures that are recognised in the art (see, for example, "Current Protocols in Molecular Biology", Ausubel *et al.* (eds). Greene Publishing Association and John Wiley Interscience, New York, 1989,1992). Probes comprising at least 15, preferably at least 30, and more preferably at  
20 least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene (SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31 and SEQ ID NO:33), are particularly useful probes.  
25 Such probes may be labelled with an analytically-detectable reagent to facilitate their identification. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes and enzymes that are capable of catalysing the formation of a detectable product. Using these probes, the ordinarily skilled artisan will be capable of isolating complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding  
30 proteins of interest from human, mammalian or other animal sources and screening such sources for related sequences, for example, for additional members of the family, type and/or subtype.

In many cases, isolated cDNA sequences will be incomplete, in that the region encoding the polypeptide will be cut short, normally at the 5' end. Several methods are available to obtain full length cDNAs, or to extend short cDNAs. Such sequences may be extended utilising a partial nucleotide sequence and employing various methods known in the art to  
5 detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed is based on the method of Rapid Amplification of cDNA Ends (RACE; see, for example, Frohman *et al.*, PNAS USA 85, 8998-9002, 1988). Recent modifications of this technique, exemplified by the Marathon<sup>TM</sup> technology (Clontech Laboratories Inc.), for example, have significantly simplified the search for longer cDNAs.  
10 A slightly different technique, termed "restriction-site" PCR, uses universal primers to retrieve unknown nucleic acid sequence adjacent a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Inverse PCR may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T. *et al.* (1988) Nucleic Acids Res. 16:8186). Another method which may be used is capture PCR which  
15 involves PCR amplification of DNA fragments adjacent a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. *et al.* (1991) PCR Methods Applic., 1, 111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J.D. *et al.* (1991); Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder<sup>TM</sup> libraries to walk genomic DNA (Clontech,  
20 Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences that contain the 5' regions of genes. Use of a randomly  
25 primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

In one embodiment of the invention, the nucleic acid molecules of the present invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is  
30 specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important step in the confirmatory correlation of those sequences with the gene-associated disease. Once a sequence has been mapped to a precise



chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that have been mapped to  
5 the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, any sequences mapping to that area may represent associated or  
10 regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to translocation, inversion, *etc.* among normal, carrier, or affected individuals.

The nucleic acid molecules of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the  
15 polypeptide in tissues by detection of the mRNAs that encode them. These techniques include *in situ* hybridization techniques and nucleotide amplification techniques, such as PCR. Results from these studies provide an indication of the normal functions of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable  
20 insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

Gene silencing approaches may also be undertaken to down-regulate endogenous expression of a gene encoding a polypeptide of the invention. RNA interference (RNAi) (Elbashir, SM *et al.*, Nature 2001, 411, 494-498) is one method of sequence specific post-  
25 transcriptional gene silencing that may be employed. Short dsRNA oligonucleotides are synthesised *in vitro* and introduced into a cell. The sequence specific binding of these dsRNA oligonucleotides triggers the degradation of target mRNA, reducing or ablating target protein expression.

Efficacy of the gene silencing approaches assessed above may be assessed through the  
30 measurement of polypeptide expression (for example, by Western blotting), and at the RNA level using TaqMan-based methodologies.

The vectors of the present invention comprise nucleic acid molecules of the invention and

may be cloning or expression vectors. The host cells of the invention, which may be transformed, transfected or transduced with the vectors of the invention may be prokaryotic or eukaryotic.

The polypeptides of the invention may be prepared in recombinant form by expression of  
5 their encoding nucleic acid molecules in vectors contained within a host cell. Such expression methods are well known to those of skill in the art and many are described in detail by Sambrook *et al.* (*supra*) and Fernandez & Hoeffler (1998, eds. "Gene expression systems. Using nature for the art of expression". Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto).

10 Generally, any system or vector that is suitable to maintain, propagate or express nucleic acid molecules to produce a polypeptide in the required host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those described in Sambrook *et al.*, (*supra*). Generally, the encoding gene can be placed under the control of a control element  
15 such as a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence encoding the desired polypeptide is transcribed into RNA in the transformed host cell.

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids,  
20 bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to  
25 deliver larger fragments of DNA than can be contained and expressed in a plasmid. The vectors pCR4-TOPO-INSP179-EC, pCR4-TOPO-INSP179-EC-SV1, pENTR\_INSP179EC-6HIS, pEAK12d\_INSP179EC-6HIS, pDEST12.2\_INSP179EC-6HIS, pENTR\_INSP179ECSV1-6HIS, pEAK12d\_INSP179ECSV1-6HIS and pDEST12.2\_INSP179ECSV1-6HIS are preferred examples of suitable vectors for use in  
30 accordance with the aspects of this invention relating to INSP179.

Particularly suitable expression systems include microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors;

yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (for example, Ti or pBR322 plasmids); or  
5 animal cell systems. Cell-free translation systems can also be employed to produce the polypeptides of the invention.

Introduction of nucleic acid molecules encoding a polypeptide of the present invention into host cells can be effected by methods described in many standard laboratory manuals, such as Davis *et al.*, Basic Methods in Molecular Biology (1986) and Sambrook *et al.*, (*supra*).  
10 Particularly suitable methods include calcium phosphate transfection, DEAE-dextran mediated transfection, transfection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection (see Sambrook *et al.*, 1989 [*supra*]; Ausubel *et al.*, 1991 [*supra*]; Spector, Goldman & Leinwald, 1998). In eukaryotic cells, expression systems may either be transient (for  
15 example, episomal) or permanent (chromosomal integration) according to the needs of the system.

The encoding nucleic acid molecule may or may not include a sequence encoding a control sequence, such as a signal peptide or leader sequence, as desired, for example, for secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into  
20 the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences can be removed by the bacterial host in post-translational processing.

In addition to control sequences, it may be desirable to add regulatory sequences that allow for regulation of the expression of the polypeptide relative to the growth of the host cell.  
25 Examples of regulatory sequences are those which cause the expression of a gene to be increased or decreased in response to a chemical or physical stimulus, including the presence of a regulatory compound or to various temperature or metabolic conditions. Regulatory sequences are those non-translated regions of the vector, such as enhancers, promoters and 5' and 3' untranslated regions. These interact with host cellular proteins to  
30 carry out transcription and translation. Such regulatory sequences may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible

promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript phagemid (Stratagene, LaJolla, CA) or pSportl™ plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the  
5 genomes of plant cells (for example, heat shock, RUBISCO and storage protein genes) or from plant viruses (for example, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence, vectors based on SV40 or EBV may be used with an  
10 appropriate selectable marker.

An expression vector is constructed so that the particular nucleic acid coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the regulatory sequences being such that the coding sequence is transcribed under the "control" of the regulatory sequences, *i.e.*,  
15 RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. In some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; *i.e.*, to maintain the reading frame.

The control sequences and other regulatory sequences may be ligated to the nucleic acid coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be  
20 cloned directly into an expression vector that already contains the control sequences and an appropriate restriction site.

For long-term, high-yield production of a recombinant polypeptide, stable expression is preferred. For example, cell lines which stably express the polypeptide of interest may be  
25 transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and  
30 recovery of cells that successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby hamster kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, HEK 293, Bowes melanoma and  
5 human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA (the "MaxBac" kit). These techniques are generally known to those skilled in the art and are  
10 described fully in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987). Particularly suitable host cells for use in this system include insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells.

There are many plant cell culture and whole plant genetic expression systems known in the art. Examples of suitable plant cellular genetic expression systems include those described  
15 in US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30, 3861-3863 (1991).

In particular, all plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be utilised, so that whole plants are recovered which contain the  
20 transferred gene. Practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and other trees, legumes and vegetables.

Examples of particularly preferred bacterial host cells include *streptococci*, *staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells.

25 Examples of particularly suitable host cells for fungal expression include yeast cells (for example, *S. cerevisiae*) and *Aspergillus* cells.

Any number of selection systems are known in the art that may be used to recover transformed cell lines. Examples include the herpes simplex virus thymidine kinase (Wigler, M. *et al.* (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I.  
30 *et al.* (1980) Cell 22:817-23) genes that can be employed in tk<sup>-</sup> or aprt<sup>±</sup> cells, respectively.

Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for

selection; for example, dihydrofolate reductase (DHFR) that confers resistance to methotrexate (Wigler, M. *et al.* (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. *et al.* (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

Although the presence or absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the relevant sequence is inserted within a marker gene sequence, transformed cells containing the appropriate sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example, fluorescence activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein (see Hampton, R. *et al.* (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. *et al.* (1983) J. Exp. Med, 158, 1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labelled hybridization or PCR probes for detecting sequences related to nucleic acid molecules encoding polypeptides of the present invention include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled polynucleotide. Alternatively, the sequences encoding the polypeptide of the invention may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesise RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides. These

procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp. (Cleveland, OH)).

Suitable reporter molecules or labels, which may be used for ease of detection, include  
5 radionucleides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Nucleic acid molecules according to the present invention may also be used to create transgenic animals, particularly rodent animals. Such transgenic animals form a further aspect of the present invention. This may be done locally by modification of somatic cells,  
10 or by germ line therapy to incorporate heritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction,  
15 anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. High performance liquid chromatography is particularly useful for purification. Well known techniques for refolding proteins may be employed to regenerate an active conformation when the polypeptide is denatured during  
20 isolation and or purification.

Specialised vector constructions may also be used to facilitate purification of proteins, as desired, by joining sequences encoding the polypeptides of the invention to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble proteins. Examples of such purification-facilitating domains include metal chelating  
25 peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the  
30 polypeptide of the invention may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an enterokinase

cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. *et al.* (1992), Prot. Exp. Purif. 3: 263-281) while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. *et al.* (1993; DNA Cell Biol. 12:441-453).

If the polypeptide is to be expressed for use in screening assays, generally it is preferred that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity techniques. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Preferred compounds are effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

Agonist or antagonist compounds may be isolated from, for example, cells, cell-free preparations, chemical libraries or natural product mixtures. These agonists or antagonists may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991).

Compounds that are most likely to be good antagonists are molecules that bind to the polypeptide of the invention without inducing the biological effects of the polypeptide upon binding to it. Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to the polypeptide of the invention and thereby inhibit or extinguish its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be inhibited, such that the normal biological activity of the polypeptide is prevented.

The polypeptide of the invention that is employed in such a screening technique may be



free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The functional response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the test compound is observed.

Methods for generating detectable signals in the types of assays described herein will be known to those of skill in the art. A particular example is cotransfecting a construct expressing a polypeptide according to the invention, or a fragment that is responsible for binding to target, in fusion with the GAL4 DNA binding domain, into a cell together with a reporter plasmid, an example of which is pFR-Luc (Stratagene Europe, Amsterdam, The Netherlands). This particular plasmid contains a synthetic promoter with five tandem repeats of GAL4 binding sites that control the expression of the luciferase gene. When a potential target or ligand is added to the cells, it will bind the GAL4-polypeptide fusion and induce transcription of the luciferase gene. The level of the luciferase expression can be monitored by its activity using a luminescence reader (see, for example, Lehman *et al.* JBC 270, 12953, 1995; Pawar *et al.* JBC, 277, 39243, 2002).

A further preferred method for identifying an agonist or antagonist of a polypeptide of the invention comprises:

- (a) contacting a labelled or unlabeled compound with the polypeptide immobilized on any solid support (for example beads, plates, matrix support, chip) and detection of the compound by measuring the label or the presence of the compound itself; or
- (b) contacting a cell expressing on the surface thereof the polypeptide, by means of artificially anchoring it to the cell membrane, or by constructing a chimeric receptor being associated with a second component capable of providing a detectable signal in response to the binding of a compound to the polypeptide, with a compound to be screened under conditions to permit binding to the polypeptide; and
- (c) determining whether the compound binds to and activates or inhibits the polypeptide

by comparing the level of a signal generated from the interaction of the compound with the polypeptide with the level of a signal in the absence of the compound.

For example, a method such as FRET detection of a ligand bound to the polypeptide in the presence of peptide co-activators (Norris *et al.*, Science 285, 744, 1999) might be used.

- 5 In further preferred embodiments, the general methods that are described above may further comprise conducting the identification of agonist or antagonist in the presence of labelled or unlabelled ligand for the polypeptide.

In another embodiment of the method for identifying agonist or antagonist of a polypeptide of the present invention comprises:

- 10 determining the inhibition of binding of a ligand to the polypeptide of the invention on any solid or cellular surface thereof, in the presence of a candidate compound under conditions to permit binding to the polypeptide, and determining the amount of ligand bound to the polypeptide. A compound capable of causing reduction of binding of a ligand is considered to be a competitor which may act as an agonist or antagonist. Preferably the ligand is  
15 labelled.

More particularly, a method of screening for a polypeptide antagonist or agonist compound comprises the steps of:

- (a) incubating a labelled ligand with a polypeptide according to the invention on any solid support or the cell surface, or a cell membrane containing a polypeptide of the invention.
- 20 (b) measuring the amount of labelled ligand bound to the polypeptide on the solid support, whole cell or the cell membrane;
- (c) adding a candidate compound to a mixture of labelled ligand and immobilized polypeptide on the solid support, the whole cell or the cell membrane of step (a) and allowing the mixture to attain equilibrium;
- 25 (d) measuring the amount of labelled ligand bound to the immobilized polypeptide or the whole cell or the cell membrane after step (c); and
- (e) comparing the difference in the labelled ligand bound in step (b) and (d), such that the compound which causes the reduction in binding in step (d) is considered to be an agonist or antagonist.
- 30 The polypeptides may be found to modulate a variety of physiological and pathological

processes in a dose-dependent manner in the above-described assays. Thus, the “functional equivalents” of the polypeptides of the invention include polypeptides that exhibit any of the same modulatory activities in the above-described assays in a dose-dependent manner. Although the degree of dose-dependent activity need not be identical to that of the polypeptides of the invention, preferably the “functional equivalents” will exhibit substantially similar dose-dependence in a given activity assay compared to the polypeptides of the invention.

In certain of the embodiments described above, simple binding assays may be used, in which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a labelled competitor. In another embodiment, competitive drug screening assays may be used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

Assays may also be designed to detect the effect of added test compounds on the production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues. The formation of binding complexes between the polypeptide and the compound being tested may then be measured.

Assay methods that are also included within the terms of the present invention are those that involve the use of the genes and polypeptides of the invention in overexpression or ablation assays. Such assays involve the manipulation of levels of these genes/polypeptides in cells and assessment of the impact of this manipulation event on the physiology of the manipulated cells. For example, such experiments reveal details of signaling and metabolic pathways in which the particular genes/polypeptides are implicated, generate information regarding the identities of polypeptides with which the studied polypeptides interact and provide clues as to methods by which related genes and proteins are regulated.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest (see

International patent application WO84/03564). In this method, large numbers of different small test compounds are synthesised on a solid substrate, which may then be reacted with the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be detected using methods  
5 that are well known in the art. Purified polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques.

The polypeptide of the invention may be used to identify membrane-bound or soluble receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a  
10 radioactive isotope, is chemically modified, or is fused to a peptide sequence that facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance and spectroscopy. Binding assays may be used for the  
15 purification and cloning of the receptor, but may also identify agonists and antagonists of the polypeptide, that compete with the binding of the polypeptide to its receptor. Standard methods for conducting screening assays are well understood in the art.

The invention also includes a screening kit useful in the methods for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, that are described above.

20 The invention includes the agonists, antagonists, ligands, receptors, substrates and enzymes, and other compounds which modulate the activity or antigenicity of the polypeptide of the invention discovered by the methods that are described above.

The invention also provides pharmaceutical compositions comprising a polypeptide, nucleic acid, ligand or compound of the invention in combination with a suitable  
25 pharmaceutical carrier. These compositions may be suitable as therapeutic or diagnostic reagents, as vaccines, or as other immunogenic compositions, as outlined in detail below.

According to the terminology used herein, a composition containing a polypeptide, nucleic acid, ligand or compound [X] is "substantially free of" impurities [herein, Y] when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least  
30 about 90% by weight of the total of X+Y in the composition, more preferably at least about 95%, 98% or even 99% by weight.

The pharmaceutical compositions should preferably comprise a therapeutically effective amount of the polypeptide, nucleic acid molecule, ligand, or compound of the invention. The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent needed to treat, ameliorate, or prevent a targeted disease or condition, or to exhibit a detectable therapeutic or preventative effect. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, for example, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

The precise effective amount for a human subject will depend upon the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. This amount can be determined by routine experimentation and is within the judgement of the clinician. Generally, an effective dose will be from 0.01 mg/kg to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

A pharmaceutical composition may also contain a pharmaceutically acceptable carrier, for administration of a therapeutic agent. Such carriers include antibodies and other polypeptides, genes and other therapeutic agents such as liposomes, provided that the carrier does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers and inactive virus particles.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may additionally contain

liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

The pharmaceutical compositions utilised in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal means. Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared.

Direct delivery of the compositions will generally be accomplished by injection, subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Dosage treatment may be a single dose schedule or a multiple dose schedule.

If the activity of the polypeptide of the invention is in excess in a particular disease state, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Preferably, such antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.

In another approach, soluble forms of the polypeptide that retain binding affinity for the ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as the use of antisense nucleic acid molecules (as described above), either internally generated or separately administered. Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. *et al.* (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Such oligonucleotides may be administered or may be generated *in situ* from expression *in vivo*.

In addition, expression of the polypeptide of the invention may be prevented by using ribozymes specific to its encoding mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, *et al.*, Curr. Opin. Struct. Biol (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified bases.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine and butosine, as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine and uridine which are not as easily recognised by endogenous endonucleases.

For treating abnormal conditions related to an under-expression of the polypeptide of the invention and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound that activates the polypeptide, *i.e.*, an agonist as described above, to alleviate the abnormal condition.

5 Alternatively, a therapeutic amount of the polypeptide in combination with a suitable pharmaceutical carrier may be administered to restore the relevant physiological balance of polypeptide.

Gene therapy may be employed to effect the endogenous production of the polypeptide by the relevant cells in the subject. Gene therapy is used to treat permanently the inappropriate  
10 production of the polypeptide by replacing a defective gene with a corrected therapeutic gene.

Gene therapy of the present invention can occur *in vivo* or *ex vivo*. *Ex vivo* gene therapy requires the isolation and purification of patient cells, the introduction of a therapeutic gene and introduction of the genetically altered cells back into the patient. In contrast, *in*  
15 *vivo* gene therapy does not require isolation and purification of a patient's cells.

The therapeutic gene is typically "packaged" for administration to a patient. Gene delivery vehicles may be non-viral, such as liposomes, or replication-deficient viruses, such as adenovirus as described by Berkner, K.L., in Curr. Top. Microbiol. Immunol., 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka, N., in Curr.  
20 Top. Microbiol. Immunol., 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For example, a nucleic acid molecule encoding a polypeptide of the invention may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide, such that the  
25 packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo* (see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

30 Another approach is the administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue.

In situations in which the polypeptides or nucleic acid molecules of the invention are



disease-causing agents, the invention provides that they can be used in vaccines to raise antibodies against the disease causing agent.

Vaccines according to the invention may either be prophylactic (*i.e.* to prevent infection) or therapeutic (*i.e.* to treat disease after infection). Such vaccines comprise immunising  
5 antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with pharmaceutically-acceptable carriers as described above, which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a  
10 bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, and other pathogens.

Since polypeptides may be broken down in the stomach, vaccines comprising polypeptides are preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration  
15 include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose  
20 containers. For example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Genetic delivery of antibodies that bind to polypeptides according to the invention may  
25 also be effected, for example, as described in International patent application WO98/55607.

The technology referred to as jet injection (see, for example, [www.powderject.com](http://www.powderject.com)) may also be useful in the formulation of vaccine compositions.

A number of suitable methods for vaccination and vaccine delivery systems are described  
30 in International patent application WO00/29428.

This invention also relates to the use of nucleic acid molecules according to the present

invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the nucleic acid molecules of the invention which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acid molecules for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki *et al.*, Nature, 324, 163-166 (1986); Bej, *et al.*, Crit. Rev. Biochem. Molec. Biol., 26, 301-334 (1991); Birkenmeyer *et al.*, J. Virol. Meth., 35, 117-126 (1991); Van Brunt, J., Bio/Technology, 8, 291-294 (1990)) prior to analysis.

In one embodiment, this aspect of the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to the invention and comparing said level of expression to a control level, wherein a level that is different to said control level is indicative of disease. The method may comprise the steps of:

- a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule of the invention and the probe;
- b) contacting a control sample with said probe under the same conditions used in step a);
- c) and detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

A further aspect of the invention comprises a diagnostic method comprising the steps of:

- a) obtaining a tissue sample from a patient being tested for disease;
- b) isolating a nucleic acid molecule according to the invention from said tissue sample; and
- c) diagnosing the patient for disease by detecting the presence of a mutation in the nucleic acid molecule which is associated with disease.

To aid the detection of nucleic acid molecules in the above-described methods, an amplification step, for example using PCR, may be included.

Deletions and insertions can be detected by a change in the size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing  
5 amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA sequences of the invention. Perfectly-matched sequences can be distinguished from mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent  
10 conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

15 Such diagnostics are particularly useful for prenatal and even neonatal testing.

Point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita *et al.*, Genomics, 5, 874-879 (1989)). For example, a sequencing primer may be used with double-stranded PCR product  
20 or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabelled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. Further, point mutations and other sequence  
25 variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA  
30 sequencing (for example, Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, Proc. Natl. Acad. Sci.

USA (1985) 85: 4397-4401).

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as microdeletions, aneuploidies, translocations, inversions, can also be detected by *in situ* analysis (see, for example, Keller *et al.*, DNA Probes, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane. Fluorescence *in situ* hybridization (FISH) is presently the most commonly applied method and numerous reviews of FISH have appeared (see, for example, Trachuck *et al.*, Science, 250, 559-562 (1990), and Trask *et al.*, Trends, Genet., 7, 149-154 (1991)).

10 In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient screening of genetic variants, mutations and polymorphisms. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic  
15 variability (see for example: M.Chee *et al.*, Science (1996), Vol 274, pp 610-613).

In one embodiment, the array is prepared and used according to the methods described in PCT application WO95/11995 (Chee *et al.*); Lockhart, D. J. *et al.* (1996) Nat. Biotech. 14: 1675-1680); and Schena, M. *et al.* (1996) Proc. Natl. Acad. Sci. 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are  
20 synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support. In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/25116 (Baldeschweiler *et al.*). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to  
25 arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including  
30 robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends itself to the efficient use of commercially-available instrumentation.

In addition to the methods discussed above, diseases may be diagnosed by methods comprising determining, from a sample derived from a subject, an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the  
5 quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Assay techniques that can be used to determine levels of a polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and are  
10 discussed in some detail above (including radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays). This aspect of the invention provides a diagnostic method which comprises the steps of: (a) contacting a ligand as described above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

15 Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex formation  
20 The amount of standard complex formation may be quantified by various methods, such as by photometric means.

Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules,  
25 ligands and other compounds of the invention. Antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently  
30 or non-covalently, with a reporter molecule. A wide variety of reporter molecules known in the art may be used, several of which are described above.

Quantities of polypeptide expressed in subject, control and disease samples from biopsied

tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to distinguish between absence, presence, and excess expression of polypeptide and to monitor regulation of polypeptide levels during therapeutic intervention. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

A diagnostic kit of the present invention may comprise:

- (a) a nucleic acid molecule of the present invention;
- (b) a polypeptide of the present invention; or
- 10 (c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to the invention; a second container containing primers useful for amplifying the nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.

In an alternative aspect of the invention, a diagnostic kit may comprise an array of nucleic acid molecules, at least one of which may be a nucleic acid molecule according to the invention.

20 To detect polypeptide according to the invention, a diagnostic kit may comprise one or more antibodies that bind to a polypeptide according to the invention; and a reagent useful for the detection of a binding reaction between the antibody and the polypeptide.

Such kits will be of use in diagnosing a disease or susceptibility to disease in which members of the leucine rich repeat containing family of proteins are implicated. Such diseases include cell proliferative disorders, autoimmune/inflammatory disorders, such as arthritis, cardiovascular disorders, neurological disorders, developmental disorders, metabolic disorders, AIDS, renal disease, infections and other pathological conditions. Such kits may also be used for the detection of reproductive disorders including infertility.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to the INSP179 polypeptides.

It will be appreciated that modification of detail may be made without departing from the scope of the invention.

### **Brief description of the Figures**

- 5 **Figure 1:** Top 10 BLASTP hits for INSP179 polypeptide sequence (SEQ ID NO:10) against NCBI-nr.

**Figure 2:** Alignment of second BLASTP hit against INSP179 polypeptide sequence (SEQ ID NO:10).

**Figure 3:** CDD domain prediction for INSP179.

- 10 **Figure 4:** Signal peptide prediction (SignalP V2.0) for INSP179 polypeptide sequence (SEQ ID NO: 10).

**Figure 5:** INSP179 DNA and polypeptide sequence. The predicted transmembrane domain is highlighted. The position and sense of the PCR primers are indicated by arrows.

- Figure 6:** INSP179-EC DNA and polypeptide sequence as cloned using primers INSP179-CP1 and INSP179-CP2. The position and sense of the PCR primers are indicated by arrows.

**Figure 7:** INSP179-EC-SV1 DNA and polypeptide sequence as cloned using primers INSP179-CP1 and INSP179-CP2. The position and sense of the PCR primers are indicated by arrows.

- 20 **Figure 8:** Alignment of INSP179-EC and INSP179-EC-SV1 nucleotide sequences.

**Figure 9:** Alignment of predicted INSP179 and the cloned INSP179-EC and INSP179-EC-SV1 polypeptide sequences.

**Figure 10:** NetNGlyco prediction output for the INSP179 polypeptide (SEQ ID NO: 10) showing five potential N-glycosylation sites above threshold score.

- 25 **Figure 11:** Results from GEO dataset at NCBI for MGC38937.

### **Examples**

#### Example 1: INSP179 Protein BLAST Results

The INSP179 polypeptide sequence (SEQ ID NO: 10) was used as a protein BLAST query sequence against the NCBI non-redundant sequence database. Figure 1 shows the top

results for the BLAST query. The second hit is for a *Mus musculus* hypothetical protein and matches INSP179 with an E value of zero. An alignment of the two proteins is given in Figure 2.

5 Example 2: Domain prediction and glycosylation sites

Figure 3 shows the CDD output for the INSP179 polypeptide sequence (SEQ ID NO: 10). The leucine rich repeat in this polypeptide is clearly visible. The INSP179 polypeptide sequence has also been shown by us to have a trans-membrane region (data not shown) using TMHMM (Krogh, A. *et al* 2001, *J Mol Biol* 305(3),567-80).

- 10 Figure 10 shows five predicted glycosylation sites for the INSP179 polypeptide at positions 45, 115, 454, 746 and 756. Such N-glycosylation is a feature of extracellular proteins.

Example 3: INSP179 signal sequence

- 15 Figure 4 shows that INSP179 is predicted to possess a signal peptide at the start of the protein. As the SignalP data in Figure 4 clearly shows, the signal peptide cleavage site is thought to be between residues 24 and 25 of the INSP179 polypeptide sequence (Nielsen, H. *et al.* 1997, *Protein Engineering*, 10, 1-6; Nielsen, H., and Krogh, A.: Prediction of signal peptides and signal anchors by a hidden Markov model. In Proceedings of the Sixth  
20 International Conference on Intelligent Systems for Molecular Biology (ISMB 6), AAAI Press, Menlo Park, California, pp. 122-130 (1998)).

Example 4: Cloning of INSP179-EC and INSP179EC-SV1

4.1 Preparation of human cDNA templates

- 25 First strand cDNA was prepared from a variety of human tissue total RNA samples (Clontech, Stratagene, Ambion, Biochain Institute and in-house preparations) using Superscript II or SuperScript III RNase H<sup>-</sup> Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol.

- For SuperScript II: Oligo (dT)<sub>15</sub> primer (1µl at 500µg/ml) (Promega), 2µg human total  
30 RNA, 1µl 10mM dNTP mix (10mM each of dATP, dGTP, dCTP and dTTP at neutral pH)



and sterile distilled water to a final volume of 12µl were combined in a 1.5ml Eppendorf tube, heated to 65°C for 5 min and chilled on ice. The contents were collected by brief centrifugation and 4µl of 5X First-Strand Buffer, 2µl 0.1 M DTT, and 1µl RnaseOUT™ Recombinant Ribonuclease Inhibitor (40 units/µl, Invitrogen) were added. The contents of the tube were mixed gently and incubated at 42°C for 2 min, then 1µl (200 units) of SuperScript II™ enzyme was added and mixed gently by pipetting. The mixture was incubated at 42°C for 50 min and then inactivated by heating at 70°C for 15 min. To remove RNA complementary to the cDNA, 1µl (2 units) of *E. coli* RNase H (Invitrogen) was added and the reaction mixture incubated at 37°C for 20 min.

- 10 For SuperScript III: 1µl Oligo(dT)<sub>20</sub> primer (50µM, Invitrogen), 2µg human total RNA, 1µl 10mM dNTP mix (10mM each of dATP, dGTP, dCTP and dTTP at neutral pH) and sterile distilled water to a final volume of 10µl were combined in a 1.5ml Eppendorf tube, heated to 65°C for 5 min and then chilled on ice. For each RT reaction a cDNA synthesis mix was prepared as follows: 2µl 10X RT buffer, 4µl 25mM MgCl<sub>2</sub>, 2µl 0.1M DTT, 1µl
- 15 RNaseOUT™ (40U/µl) and 1µl SuperScript III™ RT enzyme were combined in a separate tube and then 10µl of this mix added to the tube containing the RNA/primer mixture. The contents of the tube were mixed gently, collected by brief centrifugation, and incubated at 50°C for 50 min. The reaction was terminated by incubating at 80°C for 5 min and the reaction mixture then chilled on ice and collected by brief centrifugation. To remove RNA
- 20 complementary to the cDNA, 1µl (2 units) of *E. coli* RNase H (Invitrogen) was added and the reaction mixture incubated at 37°C for 20 min.

The final 21µl reaction mix was diluted by adding 179µl sterile water to give a total volume of 200µl. The RNA samples were combined into pools such that each pool contained five different cDNA samples. 5µl of each cDNA pool was used as a template for

25 PCR in a 50µl final reaction volume and this consisted of 1µl of each cDNA sample in that pool. This represented approximately 20ng of each individual cDNA template.

#### 4.2 Gene specific cloning primers for PCR

A pair of PCR primers having a length of between 18 and 30 bases were designed to amplify the INSP179 predicted extracellular domain cds using Primer Designer Software

30 (Scientific & Educational Software, PO Box 72045, Durham, NC 27722-2045, USA). PCR primers were optimized to have a T<sub>m</sub> close to 55 ± 10°C and a GC content of 40-60%.

Primers were selected which had high selectivity for the target sequence (INSP179) with little or no specific priming.

#### 4.3 PCR amplification of INSP179 from human cDNA templates

Gene-specific cloning primers (INSP179-CP1 and INSP179-CP2, Table 1) were designed to amplify a cDNA fragment of 1089 bp covering the full length of the INSP179 extracellular domain cds. The primer pair was used with pools of human cDNA samples and with individual cDNA templates derived from colon as PCR templates. PCR was performed in a final volume of 50µl containing 1X Platinum<sup>®</sup> *Taq* High Fidelity (HiFi) buffer, 2mM MgSO<sub>4</sub>, 200µM dNTPs, 0.2µM of each cloning primer, 1 unit of Platinum<sup>®</sup> *Taq* DNA Polymerase High Fidelity (HiFi) (Invitrogen), approximately 100ng of pool cDNA or 20ng of individual template cDNA, and either 0X or 1X PCR<sub>x</sub> Enhancer solution (Invitrogen). Cycling was performed using an MJ Research DNA Engine, programmed as follows: 94°C, 2 min; 40 cycles of 94°C, 30 sec, 55°C, 30 sec, and 68°C, 2 min; followed by 1 cycle at 68°C for 7 min and a holding cycle at 4°C.

30µl of each amplification product was visualized on a 0.8% agarose gel in 1 X TAE buffer (Invitrogen). Products of the expected molecular weight were purified from the gel using the Qiagen MinElute DNA Purification System (Qiagen), eluted in 10µl of EB buffer (10mM Tris.Cl, pH 8.5) and subcloned directly.

#### 4.4 Subcloning of PCR Products

The PCR products were subcloned into the topoisomerase I modified cloning vector (pCR4-TOPO) using the TA cloning kit purchased from the Invitrogen Corporation using the conditions specified by the manufacturer. Briefly, 4µl of gel purified PCR product was incubated for 15 min at room temperature with 1µl of TOPO vector and 1µl salt solution. The reaction mixture was then transformed into *E. coli* strain TOP10 (Invitrogen) as follows: a 50µl aliquot of One Shot TOP10 cells was thawed on ice and 2µl of TOPO reaction was added. The mixture was incubated for 15 min on ice and then heat shocked by incubation at 42°C for exactly 30s. Samples were returned to ice and 250µl of warm (room temperature) SOC media was added. Samples were incubated with shaking (220 rpm) for 1h at 37°C. The transformation mixture was then plated on L-broth (LB) plates containing ampicillin (100µg/ml) and incubated overnight at 37 °C.

#### 4.5 Plasmid DNA preparation and sequencing

A number of colonies were inoculated into 5 ml L-Broth (LB) containing ampicillin (100µg/ml) and grown overnight at 37°C with shaking at 220 rpm. Miniprep plasmid DNA was prepared from the 5 ml culture using a Biorobot 8000 robotic system (Qiagen) or Wizard Plus SV Minipreps kit (Promega cat. no. 1460) according to the manufacturer's instructions. Plasmid DNA was eluted in 80µl of sterile water. The DNA concentration was measured using an Eppendorf BO photometer or Spectramax 190 photometer (Molecular Devices). Plasmid DNA (200-500ng) was subjected to DNA sequencing with the T7 and T3 primers using the BigDye Terminator system (Applied Biosystems cat. no. 4390246) according to the manufacturer's instructions. The primer sequence is shown in Table 1. Sequencing reactions were purified using Dye-Ex columns (Qiagen) or Montage SEQ 96 cleanup plates (Millipore cat. no. LSKS09624) then analyzed on an Applied Biosystems 3700 sequencer.

Sequence analysis identified a clone, amplified from a colon cDNA template, which contained the expected INSP179 extracellular domain sequence. The sequence of the cloned cDNA fragment is shown in Figure 6. The cloned PCR product is contained in plasmid pCR4-TOPO-INSP179-EC. A second clone was identified, amplified from a colon tumour cDNA template, which was missing exons 2 and 3 of the INSP179 sequence. The sequence of this cloned cDNA fragment is shown in Figure 7. The cloned PCR product is contained in pCR4-TOPO-INSP179-EC-SV1. The nucleotide sequence alignment between the cloned INSP179 extracellular domain and the cloned INSP179 extracellular SV1 domain is shown in Figure 8. The amino acid alignment between the INSP179 predicted extracellular domain and the cloned extracellular domain and cloned extracellular SV1 domain is shown in Figure 9.

#### Example 5: Construction of mammalian cell expression vectors for INSP179EC

Plasmid pCR4-TOPO-INSP179-EC was used as PCR template to generate pEAK12d and pDEST12.2 expression clones containing the INSP179EC ORF sequence with a 3' sequence encoding a 6HIS tag using the Gateway<sup>TM</sup> cloning methodology (Invitrogen).

5.1 Generation of Gateway compatible INSP179EC ORF fused to an in frame 6HIS tag sequence.

The first stage of the Gateway cloning process involves a two step PCR reaction which generates the ORF of INSP179EC flanked at the 5' end by an attB1 recombination site and Kozak sequence, and flanked at the 3' end by a sequence encoding an in-frame 6 histidine (6HIS) tag, a stop codon and the attB2 recombination site (Gateway compatible cDNA).

- 5 The first PCR reaction (in a final volume of 50µl) contains respectively: 1µl (30ng) of plasmid pCR4-TOPO-INSP179-EC, 1.5µl dNTPs (10mM), 10µl of 10X Pfx polymerase buffer, 1µl MgSO<sub>4</sub> (50 mM), 0.5µl each of gene specific primer (100µM) (INSP179EC-EX1 and INSP179EC-EX2), and 0.5µl Platinum Pfx DNA polymerase (Invitrogen). The PCR reaction was performed using an initial denaturing step of 95°C for 2 min, followed  
10 by 12 cycles of 94°C for 15s; 55°C for 30s and 68°C for 2 min; and a holding cycle of 4°C. The amplification product was directly purified using the Wizard PCR Preps DNA Purification System (Promega) and recovered in 50µl sterile water according to the manufacturer's instructions.

- The second PCR reaction (in a final volume of 50µl) contained 10µl purified PCR1  
15 product, 1.5µl dNTPs (10mM), 5µl of 10X Pfx polymerase buffer, 1µl MgSO<sub>4</sub> (50mM), 0.5µl of each Gateway conversion primer (100µM) (GCP forward and GCP reverse) and 0.5µl of Platinum Pfx DNA polymerase. The conditions for the 2nd PCR reaction were: 95°C for 1 min; 4 cycles of 94°C, 15 sec; 50°C, 30 sec and 68°C for 2 min; 25 cycles of 94°C, 15 sec; 55°C, 30 sec and 68°C, 2 min; followed by a holding cycle of 4°C. The PCR  
20 mixture was cleaned up directly using the Wizard SV Gel and PCR Clean-up System (Promega) and recovered in 50µl sterile water according to the manufacturer's instructions. A 10µl aliquot was visualized on 0.8% agarose gel in 1 X TAE buffer (Invitrogen) in order to verify that the product was of the expected molecular weight (1089 + 70 = 1159 bp).

- 5.2 Subcloning of Gateway compatible INSP179EC ORF into Gateway entry vector  
25 pDONR221 and expression vectors pEAK12d and pDEST12.2

- The second stage of the Gateway cloning process involves subcloning of the Gateway modified PCR product into the Gateway entry vector pDONR221 as follows: 5µl of purified product from PCR2 were incubated with 1.5µl pDONR221 vector (0.1µg/µl), 2µl BP buffer and 1.5µl of BP clonase enzyme mix (Invitrogen) in a final volume of 10µl at  
30 RT for 1h. The reaction was stopped by addition of proteinase K 1µl (2µg/µl) and incubated at 37°C for a further 10 min. An aliquot of this reaction (2µl) was used to transform *E. coli* strain TOP10 (Invitrogen) as follows: a 50µl aliquot of One Shot TOP10

cells was thawed on ice and 2µl of reaction mixture added. The mixture was incubated for 30 min on ice and then heat shocked by incubation at 42°C for exactly 30 s. Samples were returned to ice and 250µl of warm SOC media (room temperature) was added. Samples were incubated with shaking (220 rpm) for 1h at 37°C. The transformation mixture was

5 then plated on L-broth (LB) plates containing kanamycin (40µg/ml) and incubated overnight at 37°C. Six of the resultant colonies were each inoculated into 1.3 ml of T-broth (TB) using a Qpix2 colony picking robot (Genetix), grown up overnight at 37°C with shaking (220 rpm), and plasmid miniprep DNA was prepared using a Qiaprep BioRobot 8000 system (Qiagen) as described above. Plasmid DNA (150-200 ng) was subjected to

10 DNA sequencing with 21M13 and M13Rev primers using the BigDyeTerminator system (Applied Biosystems cat. no. 4336919) according to the manufacturer's instructions. The primer sequences are shown in Table 1. Sequencing reactions were purified using Montage SEQ 96 cleanup plates (Millipore cat. no. LSKS09624) then analyzed on an Applied Biosystems 3700 sequencer.

15 Plasmid eluate (2µl or approx. 150ng) from one of the clones which contained the correct sequence (pENTR\_INSP179EC-6HIS) was then used in a recombination reaction containing 1.5µl of either pEAK12d vector or pDEST12.2 vector (0.1µg/µl), 2µl LR buffer and 1.5µl of LR clonase (Invitrogen) in a final volume of 10µl. The mixture was incubated at RT for 1 h, stopped by addition of proteinase K (2µg) and incubated at 37°C for a

20 further 10 min. An aliquot of this reaction (2µl) was used to transform *E. coli* strain TOP10 (Invitrogen) as follows: a 50µl aliquot of One Shot TOP10 cells was thawed on ice and 2µl of reaction mixture added. The mixture was incubated for 30 min on ice and then heat shocked by incubation at 42°C for exactly 30s. Samples were returned to ice and 250µl of warm SOC media (room temperature) was added. Samples were incubated with shaking

25 (220 rpm) for 1 h at 37°C. The transformation mixture was then plated on L-broth (LB) plates containing ampicillin (100µg/ml) and incubated overnight at 37°C.

Plasmid miniprep DNA was prepared from 5ml cultures from 6 of the resultant colonies subcloned in each vector using a Qiaprep BioRobot 8000 system (Qiagen). Plasmid DNA (200-500ng) in the pEAK12d vector was subjected to DNA sequencing with the

30 sequencing primers pEAK12F and pEAK12R, and gene-specific sequencing primers INSP179-SP1, INSP179-SP2 and INSP179-SP3, as described above. Plasmid DNA (200-500ng) in the pDEST12.2 vector was subjected to DNA sequencing with the sequencing

primers 21M13, M13Rev, INSP179-SP1, INSP179-SP2 and INSP179-SP3, as described above. Primer sequences are shown in Table 1.

CsCl gradient purified maxi-prep DNA was prepared from a 500 ml culture of the a sequence verified clone (pEAK12d\_INSP179EC-6HIS) using the method described by  
5 Sambrook J. *et al.*, 1989 (in Molecular Cloning, a Laboratory Manual, 2<sup>nd</sup> edition, Cold Spring Harbor Laboratory Press). Plasmid DNA was resuspended at a concentration of 1µg/µl in sterile water (or 10 mM Tris-HCl pH 8.5) and stored at -20°C.

Endotoxin-free maxi-prep DNA was prepared from a 500 ml culture of a sequence verified clone (pDEST12.2\_INSP179EC-6HIS) using the EndoFree Plasmid Mega kit (Qiagen)  
10 according to the manufacturer's instructions. Purified plasmid DNA was resuspended in endotoxin free TE buffer at a final concentration of at least 3µg/µl and stored at -20°C.

#### Example 6: Construction of mammalian cell expression vectors for INSP179ECSV1

Plasmid pCR4-TOPO-INSP179-EC-SV1 was used as PCR template to generate pEAK12d  
15 and pDEST12.2 expression clones containing the INSP179ECSV1 ORF sequence with a 3' sequence encoding a 6HIS tag using the Gateway<sup>TM</sup> cloning methodology (Invitrogen).

##### 6.1 Generation of Gateway compatible INSP179ECSV1 ORF fused to an in frame 6HIS tag sequence.

The first stage of the Gateway cloning process involves a two step PCR reaction which  
20 generates the ORF of INSP179-ECSV1 flanked at the 5' end by an attB1 recombination site and Kozak sequence, and flanked at the 3' end by a sequence encoding an in-frame 6 histidine (6HIS) tag, a stop codon and the attB2 recombination site (Gateway compatible cDNA).

The first PCR reaction (in a final volume of 50µl) contains respectively: 1µl (30ng) of  
25 plasmid pCR4-TOPO-INSP179-EC-SV1, 1.5µl dNTPs (10 mM), 10µl of 10X Pfx polymerase buffer, 1µl MgSO<sub>4</sub> (50mM), 0.5µl each of gene specific primer (100µM) (INSP179EC-EX1 and INSP179EC-EX2), and 0.5µl Platinum Pfx DNA polymerase (Invitrogen). The PCR reaction was performed using an initial denaturing step of 95°C for 2 min, followed by 12 cycles of 94°C for 15s; 55°C for 30s and 68°C for 2 min; and a  
30 holding cycle of 4°C. The amplification product was directly purified using the Wizard

PCR Preps DNA Purification System (Promega) and recovered in 50µl sterile water according to the manufacturer's instructions.

The second PCR reaction (in a final volume of 50µl) contained 10µl purified PCR1 product, 1.5µl dNTPs (10mM), 5µl of 10X Pfx polymerase buffer, 1µl MgSO<sub>4</sub> (50mM),  
5 0.5µl of each Gateway conversion primer (100µM) (GCP forward and GCP reverse) and 0.5µl of Platinum Pfx DNA polymerase. The conditions for the 2nd PCR reaction were: 95°C for 1 min; 4 cycles of 94°C, 15 sec; 50°C, 30 sec and 68°C for 2 min; 25 cycles of 94°C, 15 sec; 55°C, 30 sec and 68°C, 2 min; followed by a holding cycle of 4°C. The PCR mixture was cleaned up directly using the Wizard SV Gel and PCR Clean-up System  
10 (Promega) and recovered in 50µl sterile water according to the manufacturer's instructions. A 10µl aliquot was visualized on 0.8% agarose gel in 1 X TAE buffer (Invitrogen) in order to verify that the product was of the expected molecular weight (729 + 70=799 bp).

## 6.2 Subcloning of Gateway compatible INSP179ECSV1 ORF into Gateway entry vector pDONR221 and expression vectors pEAK12d and pDEST12.2

15 The second stage of the Gateway cloning process involves subcloning of the Gateway modified PCR product into the Gateway entry vector pDONR221 as follows: 5µl of cleaned product from PCR2 were incubated with 1.5µl pDONR221 vector (0.1µg/µl), 2µl BP buffer and 1.5µl of BP clonase enzyme mix (Invitrogen) in a final volume of 10µl at RT for 1h. The reaction was stopped by addition of 1µl proteinase K (2µg/µl) and  
20 incubated at 37°C for a further 10 min. An aliquot of this reaction (2µl) was used to transform *E. coli* strain TOP10 (Invitrogen) as follows: a 50µl aliquot of One Shot TOP10 cells was thawed on ice and 2µl of reaction mixture added. The mixture was incubated for 30 min on ice and then heat shocked by incubation at 42°C for exactly 30s. Samples were returned to ice and 250µl of warm SOC media (room temperature) was added. Samples  
25 were incubated with shaking (220 rpm) for 1h at 37°C. The transformation mixture was then plated on L-broth (LB) plates containing kanamycin (40µg/ml) and incubated overnight at 37°C. Six of the resultant colonies were each inoculated into 1.3 ml of T-broth (TB) using a Qpix2 colony picking robot (Genetix), grown up overnight at 37°C with shaking (220 rpm), and plasmid miniprep DNA was prepared using a Qiaprep BioRobot  
30 8000 system (Qiagen) as described above. Plasmid DNA (150-200ng) was subjected to DNA sequencing with 21M13 and M13Rev primers using the BigDyeTerminator system (Applied Biosystems cat. no. 4336919) according to the manufacturer's instructions. The

primer sequences are shown in Table 1. Sequencing reactions were purified using Montage SEQ 96 cleanup plates (Millipore cat. no. LSKS09624) then analyzed on an Applied Biosystems 3700 sequencer.

Plasmid eluate (2µl or approx. 150ng) from one of the clones which contained the correct  
5 sequence (pENTR\_INSP179ECSV1-6HIS) was then used in a recombination reaction containing 1.5µl of either pEAK12d vector or pDEST12.2 vector (0.1µg/µl), 2µl LR buffer and 1.5µl of LR clonase (Invitrogen) in a final volume of 10µl. The mixture was incubated at RT for 1 h, stopped by addition of 1µl proteinase K (2µg/µl) and incubated at 37°C for a further 10 min. An aliquot of this reaction (2µl) was used to transform *E. coli* strain TOP10  
10 (Invitrogen) as follows: a 50µl aliquot of One Shot TOP10 cells was thawed on ice and 2µl of reaction mixture added. The mixture was incubated for 30 min on ice and then heat shocked by incubation at 42°C for exactly 30 s. Samples were returned to ice and 250µl of warm SOC media (room temperature) was added. Samples were incubated with shaking (220 rpm) for 1 h at 37°C. The transformation mixture was then plated on L-broth (LB)  
15 plates containing ampicillin (100µg/ml) and incubated overnight at 37°C.

Plasmid miniprep DNA was prepared from 5ml cultures from 6 of the resultant colonies subcloned in each vector using a Qiaprep BioRobot 8000 system (Qiagen). Plasmid DNA (200-500ng) in the pEAK12d vector was subjected to DNA sequencing with the sequencing primers pEAK12F, pEAK12R, INSP179-SP1 and INSP179-SP2, as described  
20 above. Plasmid DNA (200-500ng) in the pDEST12.2 vector was subjected to DNA sequencing with the sequencing primers 21M13 and M13Rev, INSP179-SP1 and INSP179-SP3, as described above. Primer sequences are shown in Table 1.

CsCl gradient purified maxi-prep DNA was prepared from a 500ml culture of the sequence verified clone (pEAK12d\_INSP179ECSV1-6HIS) using the method described by  
25 Sambrook J. *et al.*, 1989 (in Molecular Cloning, a Laboratory Manual, 2<sup>nd</sup> edition, Cold Spring Harbor Laboratory Press). Plasmid DNA was resuspended at a concentration of 1µg/µl in sterile water (or 10mM Tris-HCl pH 8.5) and stored at -20°C.

Endotoxin-free maxi-prep DNA was prepared from a 500 ml culture of the sequence verified clone (pDEST12.2\_INSP179ECSV1-6HIS) using the EndoFree Plasmid Mega kit  
30 (Qiagen) according to the manufacturer's instructions. Purified plasmid DNA was resuspended in endotoxin free TE buffer at a final concentration of at least 3µg/µl and stored at -20°C.



Table 1

Primer	Sequence (5'-3')
INSP179-CP1	ATG AAA AAC CTC TAT TTC AGA GTC ATT AC (SEQ ID NO: 36)
INSP179-CP2	CTG CAC ATC GCG GGT GCT TCT AAC (SEQ ID NO: 37)
INSP179EC-EX1	GCA GGC TTC <u>GCC ACC</u> ATG AAA AAC CTC TAT TTC AG (SEQ ID NO: 38)
INSP179EC-EX2	<i>TG ATG GTG ATG GTG</i> CTG CAC ATC GCG GGT GCT TCT (SEQ ID NO: 39)
INSP179-SP1	ATG AAA AAC CTC TAT TTC AGA GTC ATT AC (SEQ ID NO: 40)
INSP179-SP2	CAA GCC TTC AAG GAC CTC AA (SEQ ID NO: 41)
INSP179-SP3	CAT GCG ATG CAG ATG AAT GG (SEQ ID NO: 42)
GCP Forward	G GGG ACA AGT TTG TAC AAA AAA GCA GGC TTC <u>GCC ACC</u> (SEQ ID NO: 43)
GCP Reverse	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT <b>TCA</b> <i>ATG GTG ATG GTG ATG GTG</i> (SEQ ID NO: 44)
pEAK12F	GCC AGC TTG GCA CTT GAT GT (SEQ ID NO: 45)
pEAK12R	GAT GGA GGT GGA CGT GTC AG (SEQ ID NO: 46)
21M13	TGT AAA ACG ACG GCC AGT (SEQ ID NO: 47)
M13REV	CAG GAA ACA GCT ATG ACC (SEQ ID NO: 48)
T7	TAA TAC GAC TCA CTA TAG G (SEQ ID NO: 49)
T3	ATT AAC CCT CAC TAA AGG (SEQ ID NO: 50)

Underlined sequence = Kozak sequence

5 **Bold** = Stop codon

*Italic* sequence = His tag

### Example 7: Role of INSP179

Results from the GEO dataset at NCBI for MGC38937 (which shares high homology with  
 10 INSP179-see example 1, and Figures 1 and 2) shows that this protein is present in a  
 number of T-cell derived cell/monocytic lines and in particular in THP-1 cells where it is  
 up-regulated by LPS treatment of the THP-1 cells over a time-course experiment (Figure  
 11). This indicates that INSP179 may be involved in the pro-inflammatory response  
 induced by bacterial lipopolysaccharide (LPS, which is typically obtained from *E.coli* and  
 15 is often used to induce inflammatory responses in T cells and macrophages).

To assay for this activity, THP-1 cells may be grown in RPMI 1640 culture medium supplemented with 10% fetal calf serum (FCS) (Gibco), 2 mM glutamine, penicillin (100 IU/ml) and streptomycin (50 µg/ml) at 37°C in humidified air with 5% CO<sub>2</sub>.

The cells are harvested and resuspended in RPMI and then treated with LPS (1-10µg/ml)  
5 (Song & Phelps, 2000, Am.J.Physiol.Lung.Cell.Mol.Physiol. 279:L110-L117). Samples can then be taken at 0, 2, 4, 6, 12 and 24 hours and the level of expression of INSP179 is determined by ELISA. The cells may also be simultaneously treated with potential agonists or antagonists to determine their effect on INSP179 expression.